

Medidas de Evaluación

Alba Sevilla Navarro

2025-09-22

Obtención de medidas de evaluación para los datasets:

- AmphibiansAndReptilesCesar Capinha de César Capinha
- USGS
- InvasiveNonNativeSpeciesInBrazil_RafaelZenni

Abrimos los datasets obtenidos con la función “Medidas_Evaluación.R”

```
library(readxl)
setwd("../Validation")
amphibians_dataset <- read_excel("AMPHIBIANSCESAR_DATASET_HABITAT.xlsx")
usgs_dataset <- read_excel("USGS_DATASET_HABITAT.xlsx")
InvasiveNonNativeSpeciesInBrazil_RafaelZenni_dataset <- read_excel("ZENNI_DATASET_HABITAT.xlsx")
```

La columna con los valores predichos será la llamada `Binary Habitat Predcited` donde:

$$\begin{cases} Y = 0 & (\text{No Freshwater}) \\ Y = 1 & (\text{Freshwater}) \end{cases}$$

La columna con los valores reales será la llamada `Binary Habitat original` donde Y=0 e Y=1 tiene la misma asignación que en el caso de los valores predichos.

$$\begin{cases} Y = 0 & (\text{No Freshwater}) \\ Y = 1 & (\text{Freshwater}) \end{cases}$$

Para comprobar la bondad del ajuste de la función que obtiene los habitats de las especies en base a la Database GBIF, a través de su paquete ‘rgbif’, obtenemos las siguientes matrices de confusión y medidas de bondad de ajuste, mediante la función `confusionMatrix` del paquete `caret` :

Table 1: ConfusionMatrix

		Original	
Predicción		No Freshwater	Freshwater
No Freshwater	A	B	
	C	D	

Con las fórmulas:

$$Sensibilidad : \frac{A}{A + C}$$

$$Especificidad : \frac{D}{B + D}$$

$$Prevalencia : \frac{A + C}{A + B + C + D}$$

$$PPV = \frac{Sensibilidad * Prevalencia}{(Sensibilidad * Prevalencia) + ((1 - Especificidad) * (1 - Prevalencia))}$$

$$NPV = \frac{Especificidad * (1 - Prevalencia)}{((1 - Sensibilidad) * Prevalencia) + (Especificidad * (1 - Prevalencia))}$$

$$DetectionRate = \frac{A}{A + B + C + D}$$

$$DetectionPrevalence = \frac{A + B}{A + B + C + D}$$

$$BalancedAccuracy = \frac{(Sensibilidad + Especificidad)}{2}$$

$$Precision = \frac{A}{A + B}$$

$$Recall = \frac{A}{A + C}$$

$$F1 = \frac{(1 + \beta^2) * Precision * Recall}{(\beta^2 * Precision) + Recall}; \quad \text{donde } \beta = 1$$

- AmphibiansAndReptilesCesar Capinha

```
# Extraer la matriz
amphibians_dataset$`1_Cesar` <- factor(amphibians_dataset$`1_Cesar`, levels = c("FRESHWATER", "NO FRESHWATER"))
amphibians_dataset$Habitat <- factor(amphibians_dataset$Habitat, levels = c("FRESHWATER", "NO FRESHWATER"))
matriz <- caret::confusionMatrix(factor(amphibians_dataset$Habitat), factor(amphibians_dataset$`1_Cesar`))

# Obtener los valores específicos de TP, FP, TN, FN
tp <- matriz["FRESHWATER", "FRESHWATER"] # Verdaderos positivos
fp <- matriz["FRESHWATER", "NO FRESHWATER"] # Falsos positivos
tn <- matriz["NO FRESHWATER", "NO FRESHWATER"] # Verdaderos negativos
fn <- matriz["NO FRESHWATER", "FRESHWATER"] # Falsos negativos

# Imprimir los resultados
cat("Verdaderos Positivos (TP):", tp, "\n")

## Verdaderos Positivos (TP): 73

cat("Falsos Positivos (FP):", fp, "\n")

## Falsos Positivos (FP): 7

cat("Verdaderos Negativos (TN):", tn, "\n")

## Verdaderos Negativos (TN): 162
```

```

cat("Falsos Negativos (FN):", fn, "\n")

## Falsos Negativos (FN): 34

# Tasa de Verdaderos Positivos (Sensibilidad o TPR)
tpr <- tp / (tp + fn)
# Tasa de Falsos Positivos (FPR)
fpr <- fp / (fp + tn)
# Tasa de Verdaderos Negativos (Especificidad o TNR)
tnr <- tn / (tn + fp)
# Tasa de Falsos Negativos (FNR)
fnr <- fn / (fn + tp)

# Mostrar los resultados
cat("Tasa de Verdaderos Positivos (TPR o Sensibilidad):", tpr, "\n")

## Tasa de Verdaderos Positivos (TPR o Sensibilidad): 0.682243

cat("Tasa de Falsos Positivos (FPR):", fpr, "\n")

## Tasa de Falsos Positivos (FPR): 0.04142012

cat("Tasa de Verdaderos Negativos (TNR o Especificidad):", tnr, "\n")

## Tasa de Verdaderos Negativos (TNR o Especificidad): 0.9585799

cat("Tasa de Falsos Negativos (FNR):", fnr, "\n")

## Tasa de Falsos Negativos (FNR): 0.317757

• USGS

# Extraer la matriz
usgs_dataset$Native.Habitat <- factor(usgs_dataset$Native.Habitat, levels = c("FRESHWATER", "NO FRESHWATER"))
usgs_dataset$Habitat <- factor(usgs_dataset$Habitat, levels = c("FRESHWATER", "NO FRESHWATER"))
matriz <- caret::confusionMatrix(factor(usgs_dataset$Habitat), factor(usgs_dataset$Native.Habitat), pos = "NO FRESHWATER")

# Obtener los valores específicos de TP, FP, TN, FN
tp <- matriz["FRESHWATER", "FRESHWATER"] # Verdaderos positivos
fp <- matriz["FRESHWATER", "NO FRESHWATER"] # Falsos positivos
tn <- matriz["NO FRESHWATER", "NO FRESHWATER"] # Verdaderos negativos
fn <- matriz["NO FRESHWATER", "FRESHWATER"] # Falsos negativos

# Imprimir los resultados
cat("Verdaderos Positivos (TP):", tp, "\n")

## Verdaderos Positivos (TP): 917

```

```

cat("Falsos Positivos (FP):", fp, "\n")

## Falsos Positivos (FP): 38

cat("Verdaderos Negativos (TN):", tn, "\n")

## Verdaderos Negativos (TN): 156

cat("Falsos Negativos (FN):", fn, "\n")

## Falsos Negativos (FN): 348

# Tasa de Verdaderos Positivos (Sensibilidad o TPR)
tpr <- tp / (tp + fn)
# Tasa de Falsos Positivos (FPR)
fpr <- fp / (fp + tn)
# Tasa de Verdaderos Negativos (Especificidad o TNR)
tnr <- tn / (tn + fp)
# Tasa de Falsos Negativos (FNR)
fnr <- fn / (fn + tp)

# Mostrar los resultados
cat("Tasa de Verdaderos Positivos (TPR o Sensibilidad):", tpr, "\n")

## Tasa de Verdaderos Positivos (TPR o Sensibilidad): 0.7249012

cat("Tasa de Falsos Positivos (FPR):", fpr, "\n")

## Tasa de Falsos Positivos (FPR): 0.1958763

cat("Tasa de Verdaderos Negativos (TNR o Especificidad):", tnr, "\n")

## Tasa de Verdaderos Negativos (TNR o Especificidad): 0.8041237

cat("Tasa de Falsos Negativos (FNR):", fnr, "\n")

## Tasa de Falsos Negativos (FNR): 0.2750988

• InvasiveNonNativeSpeciesInBrazil_RafaelZenni

# Extraer la matriz
InvasiveNonNativeSpeciesInBrazil_RafaelZenni_dataset$Habitat_Original <- factor(InvasiveNonNativeSpeciesInBrazil_RafaelZenni_dataset$Habitat_Original)
InvasiveNonNativeSpeciesInBrazil_RafaelZenni_dataset$Habitat <- factor(InvasiveNonNativeSpeciesInBrazil_RafaelZenni_dataset$Habitat)
matriz <- caret::confusionMatrix(factor(InvasiveNonNativeSpeciesInBrazil_RafaelZenni_dataset$Habitat), ...)

# Obtener los valores específicos de TP, FP, TN, FN
tp <- matriz["FRESHWATER", "FRESHWATER"] # Verdaderos positivos

```

```

fp <- matriz["FRESHWATER", "NO FRESHWATER"] # Falsos positivos
tn <- matriz["NO FRESHWATER", "NO FRESHWATER"] # Verdaderos negativos
fn <- matriz["NO FRESHWATER", "FRESHWATER"] # Falsos negativos

# Imprimir los resultados
cat("Verdaderos Positivos (TP):", tp, "\n")

## Verdaderos Positivos (TP): 130

cat("Falsos Positivos (FP):", fp, "\n")

## Falsos Positivos (FP): 36

cat("Verdaderos Negativos (TN):", tn, "\n")

## Verdaderos Negativos (TN): 270

cat("Falsos Negativos (FN):", fn, "\n")

## Falsos Negativos (FN): 8

# Tasa de Verdaderos Positivos (Sensibilidad o TPR)
tpr <- tp / (tp + fn)
# Tasa de Falsos Positivos (FPR)
fpr <- fp / (fp + tn)
# Tasa de Verdaderos Negativos (Especificidad o TNR)
tnr <- tn / (tn + fp)
# Tasa de Falsos Negativos (FNR)
fnr <- fn / (fn + tp)

# Mostrar los resultados
cat("Tasa de Verdaderos Positivos (TPR o Sensibilidad):", tpr, "\n")

## Tasa de Verdaderos Positivos (TPR o Sensibilidad): 0.942029

cat("Tasa de Falsos Positivos (FPR):", fpr, "\n")

## Tasa de Falsos Positivos (FPR): 0.1176471

cat("Tasa de Verdaderos Negativos (TNR o Especificidad):", tnr, "\n")

## Tasa de Verdaderos Negativos (TNR o Especificidad): 0.8823529

cat("Tasa de Falsos Negativos (FNR):", fnr, "\n")

## Tasa de Falsos Negativos (FNR): 0.05797101

```

En resumen:

```

calcular_metricas <- function(pred, real) {
  cm <- caret::confusionMatrix(factor(pred), factor(real), positive = "FRESHWATER")
  data.frame(
    BalancedAccuracy = cm$byClass["Balanced Accuracy"],
    Sensitivity = cm$byClass["Sensitivity"],
    Specificity = cm$byClass["Specificity"],
    Pos_Pred_Value = cm$byClass["Pos Pred Value"],
    Neg_Pred_Value = cm$byClass["Neg Pred Value"]
  )
}

# Calcular métricas para cada base
amphibians_metrics <- calcular_metricas(amphibians_dataset$`1_Cesar`, amphibians_dataset$Habitat)
usgs_metrics <- calcular_metricas(usgs_dataset$Habitat, usgs_dataset$Native.Habitat)
InvasiveNonNativeSpeciesInBrazil_RafaelZenni_metrics <- calcular_metricas(InvasiveNonNativeSpeciesInBrazil_RafaelZenni)

library(tidyjson)
# Juntar en una sola tabla
tabla_final <- bind_rows(
  AmphibiansAndReptiles_CesarCapinha = amphibians_metrics,
  USGS = usgs_metrics,
  InvasiveNonNativeSpeciesInBrazil_RafaelZenni = InvasiveNonNativeSpeciesInBrazil_RafaelZenni_metrics,
  .id = "Database"
)
tabla_final <- as.data.frame(tabla_final)
tabla_final

##                                         Database BalancedAccuracy
## Balanced Accuracy...1      AmphibiansAndReptiles_CesarCapinha      0.8695153
## Balanced Accuracy...2                  USGS          0.7645124
## Balanced Accuracy...3 InvasiveNonNativeSpeciesInBrazil_RafaelZenni      0.9121910
##                               Sensitivity Specificity Pos_Pred_Value Neg_Pred_Value
## Balanced Accuracy...1      0.9125000     0.8265306      0.6822430      0.9585799
## Balanced Accuracy...2      0.7249012     0.8041237      0.9602094      0.3095238
## Balanced Accuracy...3      0.9420290     0.8823529      0.7831325      0.9712230

```

MATRICES DE CONFUSION:

```

# Instalar paquetes si no están instalados
if(!require(caret)) install.packages("caret")
if(!require(ggplot2)) install.packages("ggplot2")
if(!require(reshape2)) install.packages("reshape2")

# Cargar las librerías
library(caret)
library(ggplot2)
library(reshape2)

# AmphibiansAndReptilesCesar Capinha
amphibians_dataset$`1_Cesar` <- factor(amphibians_dataset$`1_Cesar`, levels = c("FRESHWATER", "NO FRESHWATER"))
amphibians_dataset$Habitat <- factor(amphibians_dataset$Habitat, levels = c("FRESHWATER", "NO FRESHWATER"))
matriz_amphibios <- caret::confusionMatrix(factor(amphibians_dataset$Habitat), factor(amphibians_dataset$`1_Cesar`))

```

```

# USGS
usgs_dataset$Native.Habitat <- factor(usgs_dataset$Native.Habitat, levels = c("FRESHWATER", "NO FRESHWATER"))
usgs_dataset$Habitat <- factor(usgs_dataset$Habitat, levels = c("FRESHWATER", "NO FRESHWATER"))
matriz_usgs <- caret::confusionMatrix(factor(usgs_dataset$Habitat), factor(usgs_dataset$Native.Habitat))

# InvasiveNonNativeSpeciesInBrazil_RafaelZenni
InvasiveNonNativeSpeciesInBrazil_RafaelZenni_dataset$Habitat_Original <- factor(InvasiveNonNativeSpeciesInBrazil_RafaelZenni_dataset$Habitat)
InvasiveNonNativeSpeciesInBrazil_RafaelZenni_dataset$Habitat <- factor(InvasiveNonNativeSpeciesInBrazil_RafaelZenni_dataset$Habitat_Original)
matriz_InvasiveNonNativeSpeciesInBrazil_RafaelZenni <- caret::confusionMatrix(factor(InvasiveNonNativeSpeciesInBrazil_RafaelZenni_dataset$Habitat), factor(InvasiveNonNativeSpeciesInBrazil_RafaelZenni_dataset$Habitat_Original))

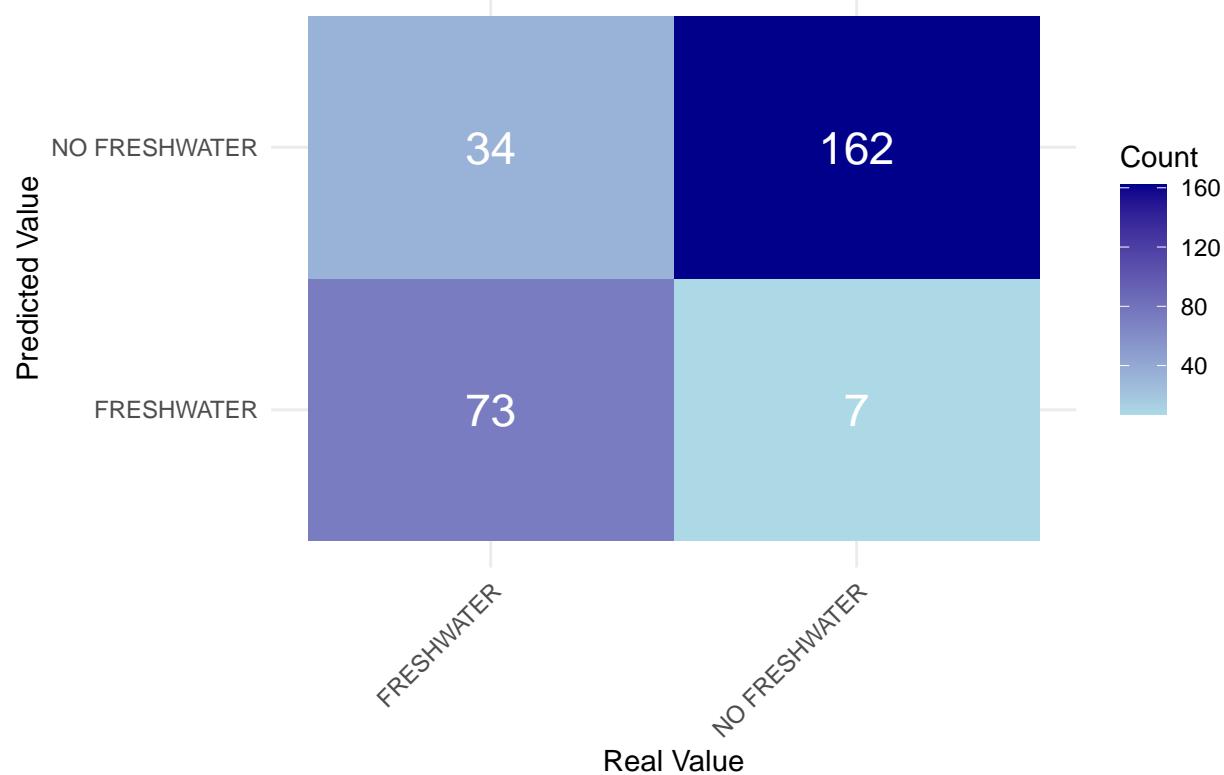
# Función para graficar la ConfusionMatrix con los números y un gradiente de azul claro a azul oscuro
plot_confusion_matrix <- function(matriz, title) {
  matriz_melt <- melt(matriz)
  colnames(matriz_melt) <- c("Predicted", "Actual", "Count")

  ggplot(matriz_melt, aes(x = Actual, y = Predicted, fill = Count)) +
    geom_tile() +
    scale_fill_gradient(low = "lightblue", high = "darkblue") + # Gradiente de colores
    geom_text(aes(label = Count), color = "white", size = 6) + # Añadir los números en las celdas
    theme_minimal() +
    labs(title = title, x = "Real Value", y = "Predicted Value") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) # Ajuste de texto en los ejes
}

# Graficar las matrices de confusión para cada conjunto de datos
plot_confusion_matrix(matriz_amphibios, "ConfusionMatrix: AmphibiansAndReptilesCesar Capinha")

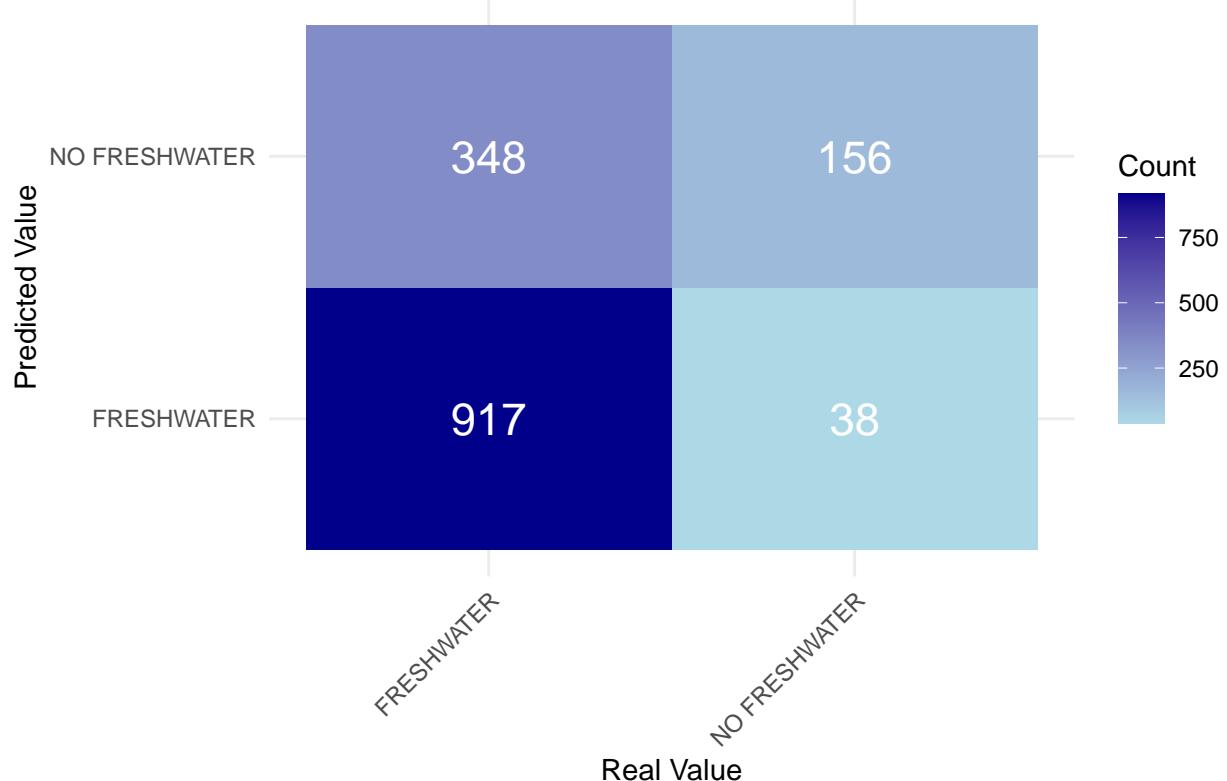
```

ConfusionMatrix: AmphibiansAndReptilesCesar Capinha



```
plot_confusion_matrix(matriz_usgs, "ConfusionMatrix: USGS")
```

ConfusionMatrix: USGS



```
plot_confusion_matrix(matrix_InvasiveNonNativeSpeciesInBrazil_RafaelZenni, "ConfusionMatrix: InvasiveNonNativeSpeciesInBrazil_RafaelZenni")
```

ConfusionMatrix: InvasiveNonNativeSpeciesInBrazil_RafaelZen

